Information Leakage and Side-Channel Attacks

Fuites d’information et attaques par canaux cachés

Collège de France, Paris, 2 mars 2022

Olivier Rioul

Télécom Paris, Institut Polytechnique de Paris, France

<olivier.rioul@telecom-paris.fr>

Shannon's Entropy: Operational Definition

Message $\underline{x} = (x_1, x_2, \ldots, x_n)$ a very long i.i.d. sequence of symbols $\sim p(x)$:

$$p(\underline{x}) = p(x_1)p(x_2)\cdots p(x_n)$$

- rearrange terms according to the number $n(x)$ of symbols equal to x:

$$p(\underline{x}) = \prod_x p(x)^{n(x)}$$

- asymptotically as $n \to +\infty$, by the law of large numbers, $\frac{n(x)}{n} \approx p(x)$:

$$p(\underline{x}) \approx \prod_x p(x)^{n \cdot p(x)} = \exp\bigl(-n \underbrace{\sum_x p(x) \log \frac{1}{p(x)}}_{\text{entropy } H = H(p)}\bigr)$$

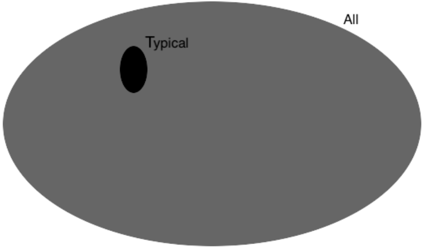
Theorem (Asymptotic Equipartition Property (AEP))

For any typical sequence, $\boxed{p(\underline{x}) \approx e^{-nH}}$ *where* $\mathbb{P}(\text{typical}) \approx 1.$

Shannon's Source Coding Theorem

To encode messages $\underline{x} = (x_1, x_2, \ldots, x_n)$ reliably:

- since $\mathbb{P}(\text{typical}) \approx 1$ it is enough to encode the N *typical* sequences ($\mathbb{P}_e \approx 0$)
- but $\mathbb{P}(\text{typical}) \approx N e^{-nH} \approx 1$, so there are about $\boxed{N \approx e^{nH}}$ typical sequences.



- coding rate (information units per symbol) : $R = \dfrac{\log N}{n} \approx H$.

Theorem (Shannon's 1st Coding Theorem)

Entropy H is an achievable (lossless) compression rate of source X

Relative Entropy (Divergence): Operational Definition

Suppose $\underline{x} = (x_1, x_2, \ldots, x_n)$ i.i.d. $\sim q(x) \neq p(x) \approx \frac{n(x)}{n}$:

$$q(\underline{x}) = q(x_1)q(x_2)\cdots q(x_n) = \prod_x q(x)^{n(x)} \approx \prod_x q(x)^{n\cdot p(x)} = \exp\Big(-n \underbrace{\sum_x p(x) \log \frac{1}{q(x)}}_{\text{cross-entropy } H(p\|q)}\Big)$$

- probability that $\underline{x} \sim q$ is p-typical ($N \approx e^{nH(p)}$ typical sequences)

$$\mathbb{P}(\text{typical}) = Ne^{-nH(p\|q)} \approx e^{nH(p)-nH(p\|q)} = \exp\Big(-n \underbrace{\sum_x p(x) \log \frac{p(x)}{q(x)}}_{\text{relative entropy } D(p\|q) = H(p\|q) - H(p) > 0}\Big)$$

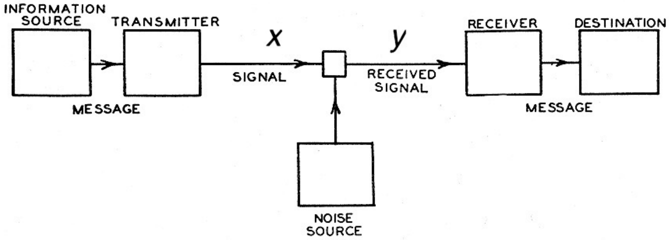
(Kullback-Leibler divergence)

Theorem (Large Deviation Bound (LDB))

Divergence $D(p\|q)$ is the deviation exponent $\boxed{\mathbb{P}(\textit{typical}) \leq e^{-nD(p\|q)}} \xrightarrow{\textit{exp.}} 0$ *if $p \neq q$*

Shannon's Channel Coding Theorem

Transmit reliably codeword $\underline{x} = (x_1, x_2, \ldots, x_n)$ in a channel $\underline{x} \to \underline{y} = (y_1, y_2, \ldots, y_n)$



- decode $\hat{\underline{x}}$ from received $\underline{y}$ with error probability $\mathbb{P}_e = \mathbb{P}(\hat{\underline{x}} \neq \underline{x})$
- "random coding" evaluate $\mathbb{P}_e$ averaged over *all possible codes* as if codewords $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N$ were drawn i.i.d. $\sim p(\underline{x})$
- typical decoding : decode $\hat{\underline{x}}$ if it is the *only* codeword jointly *typical* with received sequence $\underline{y}$, i.e., $(\underline{x}, \underline{y})$ is typical for $p(x, y)$.

Shannon's Channel Coding Theorem

- another (independent) codeword $\underline{x}' \sim q(x', y) = p(x')p(y)$ can be also jointly typical with $\underline{y}$ (according to $p(x, y)$) with probability

$$e^{-nD(p\|q)} = \exp\Bigl(-n \underbrace{\sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}}_{\text{mutual information } I(X;Y)}\Bigr) = e^{-nI(X;Y)}$$

- coding rate (information units per symbol) : $R = \dfrac{\log N}{n}$ for N codewords:

$$\mathbb{P}_e \approx (N-1)e^{-nI(X;Y)} \approx e^{n(R-I(X;Y))} \xrightarrow{\text{exp.}} 0$$

when $\boxed{R < I(X;Y)}$ (maximized for some optimal choice of $p(x)$).

Theorem (Shannon's 2nd Coding Theorem)

Any $R <$ Capacity $C = \max\limits_{p(x)} I(X;Y)$ can be a reliable transmission rate over channel $X \to Y$.

Achievability and Converse Results

- previous results are
  - *achievability* results: what one can actually do to approach a limit (H or $C\ldots$).
  - asymptotic (as $n \to +\infty$)
- converse theorems establish limits that cannot be exceeded, e.g., using

Data Processing Inequality (DPI) "processing can only decrease information"

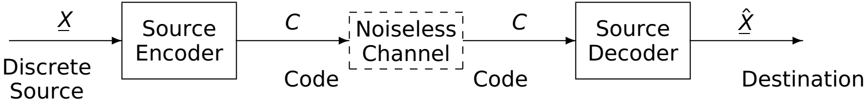
If $W - X - Y - Z$ is a Markov chain, $\boxed{I(X;Y) \geq I(W;Z)}$

Fano's Inequality [Fano'52]

$X - Y - \hat{X} = \hat{x}(Y)$ with M-ary equiprobable X, success probability $\mathbb{P}_s = \mathbb{P}(\hat{X} = X) = 1 - \mathbb{P}_e$

$$H(X|Y) \leq h(\mathbb{P}_e) + \mathbb{P}_e \log(M-1) \qquad \Longleftrightarrow \qquad I(X;Y) \geq d(\mathbb{P}_s \| \tfrac{1}{M})$$

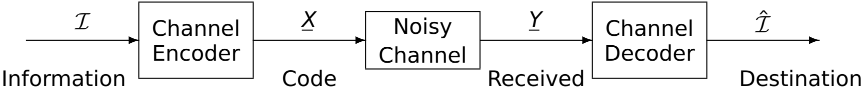
Converse Coding Theorems



By the data processing inequality, $I(\underline{X};\hat{\underline{X}}) \leq I(C;C) = H(C) \leq \log M$.

By Fano's inequality, $I(\underline{X};\hat{\underline{X}}) = H(\underline{X}) - H(X|\hat{\underline{X}}) \approx H(\underline{X}) = nH$ if $\mathbb{P}_e \approx 0$

Thus if $\mathbb{P}_e \approx 0$, $\boxed{R \geq H}$: entropy H is the optimal bound.



By the data processing inequality, $nC \geq nI(X;Y) \geq I(\underline{X};\underline{Y}) \geq I(\mathcal{I};\hat{\mathcal{I}})$

By Fano's inequality, $I(\mathcal{I};\hat{\mathcal{I}}) \geq d(\mathbb{P}_s \| 1/M) \approx \log M$ if $\mathbb{P}_s \approx 1$.

Thus if $\mathbb{P}_e \approx 0$, $\boxed{R \leq C}$: capacity C is the optimal bound.

Parametric Estimation

Observed data $\underline{x} = (x_1, x_2, \ldots, x_n)$ be a very long i.i.d. sequence $\sim p_{\theta^*} \ll \mu$.
Model $\theta \mapsto p_\theta$ is known:

$$p_\theta(\underline{x}) = p_\theta(x_1)p_\theta(x_2)\cdots p_\theta(x_n)$$

Find an asymptotically optimal estimator $\hat{\theta}(\underline{x})$ of θ^*.

- taking logarithms, asymptotically as $n \to +\infty$, by the law of large numbers,

$$\frac{1}{n}\log p_\theta(\underline{x}) = \frac{1}{n}\sum_1^n \log p_\theta(x_i) \to \mathbb{E}_{\theta^*}\log p_\theta(X) = -H(p_{\theta^*}\|p_\theta)$$

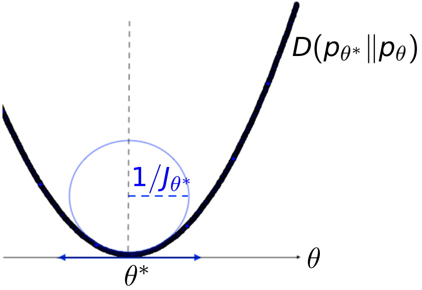
- divergence $D(p_{\theta^*}\|p_\theta) = H(p_{\theta^*}\|p_\theta) - H(p_{\theta^*}) \geq 0$ is minimum $= 0$ iff $p_\theta = p^*_\theta$.
(i.e., by identifiability $\theta = \theta^*$)
- asymptotically as $n \to +\infty$,

$$\theta^* = \arg\min_\theta D(p_{\theta^*}\|p_\theta) \iff \hat{\theta}(\underline{x}) = \arg\max_\theta \frac{1}{n}\log p_\theta(\underline{x}) \text{ (maximum likelihood)}$$

Fisher's Information: Operational Definition

At the minimum $\theta = \theta^*$:

- null gradient $\frac{\partial}{\partial\theta} D(p_{\theta^*} \| p_\theta)\big|_{\theta=\theta^*} = -\mathbb{E} S_\theta(X) = 0$
  where score $S_\theta(X) = \frac{\partial}{\partial\theta} \log p_\theta(X)$.
- curvature $J_{\theta^*} = \frac{\partial^2}{\partial\theta^2} D(p_{\theta^*} \| p_\theta)\big|_{\theta=\theta^*} \geq 0$
  (Fisher information)

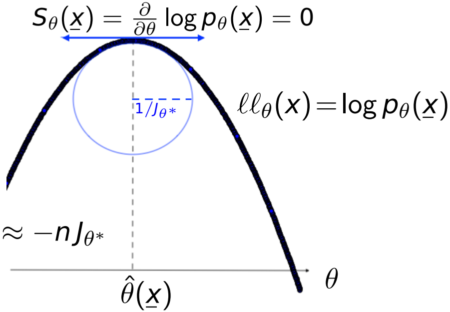


$$J_\theta = -\int \frac{\partial^2}{\partial\theta^2}\Big(\log p_\theta(x)\Big) p_\theta(x)\, \mathrm{d}\mu(x) = \underbrace{\int \Big(\frac{\partial}{\partial\theta} \log p_\theta(x)\Big)^2 p_\theta(x)\, \mathrm{d}\mu(x)}_{\mathrm{Var}(S_\theta(X))}$$

Fisher's Information: Operational Definition

Therefore, the maximum likelihood estimator $\hat\theta(\underline{x}) = \arg\max_\theta \log p_\theta(\underline{x})$ satisfies:

- asymptotically, $0 = \frac{1}{n} S_\theta(\underline{x}) \approx \mathbb{E}(S_\theta(X))$ at $\theta = \hat\theta$
  hence $D(p_{\theta^*} \| p_{\hat\theta}) \to 0$ and $\boxed{\hat\theta \to \theta^*}$ as $n \to \infty$;
- asymptotically, $\frac{S_{\theta^*}(\underline{x}) - \overbrace{S_{\hat\theta}(\underline{x})}^{=0}}{\theta^* - \hat\theta} \to \frac{\partial^2}{\partial\theta^2} \log p_\theta(\underline{x})\big|_{\theta=\theta^*} \approx -nJ_{\theta^*}$
  hence $\mathbb{E}(\hat\theta - \theta^*)^2 \sim \frac{nJ_{\theta^*}}{n^2 J_{\theta^*}^2} = \frac{1}{nJ_{\theta^*}}$, i.e.,



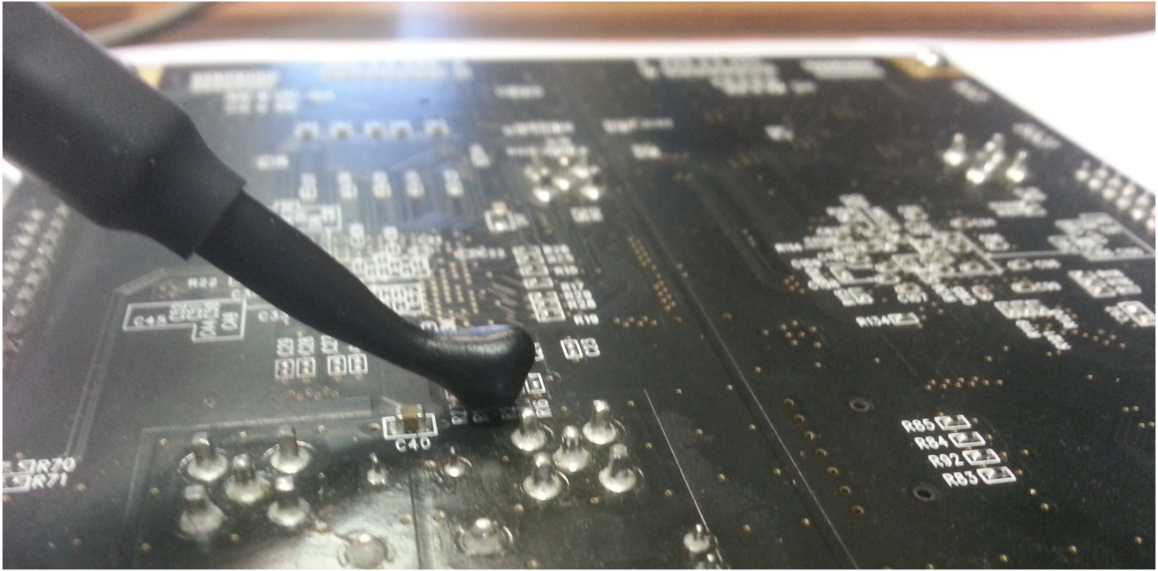
Theorem (Fisher's Estimation Theorem)

Asymptotically, ML estimation has MSE $\approx \frac{1}{nJ_{\theta^*}}$ *for observation* X

Converse theorem: Cramér-Rao bound (Fréchet-Darmois, 1943)

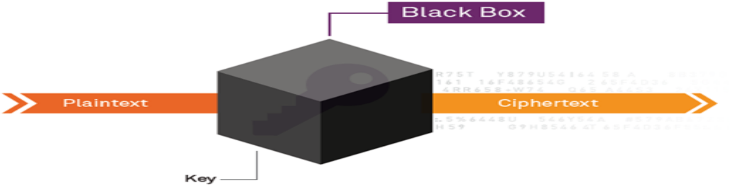
If $\hat\theta$ is unbiased, $\text{MSE} = \mathrm{Var}(\hat\theta) \geq \frac{1}{nJ_{\theta^*}}$

Information Leakage: Side-Channel Analysis

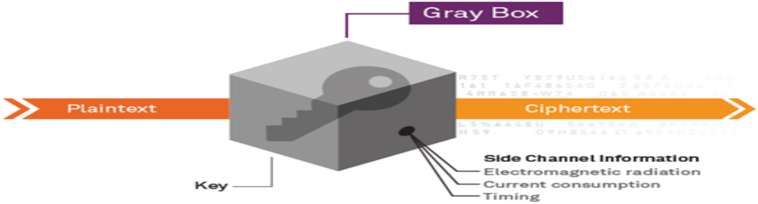


Physical Problem

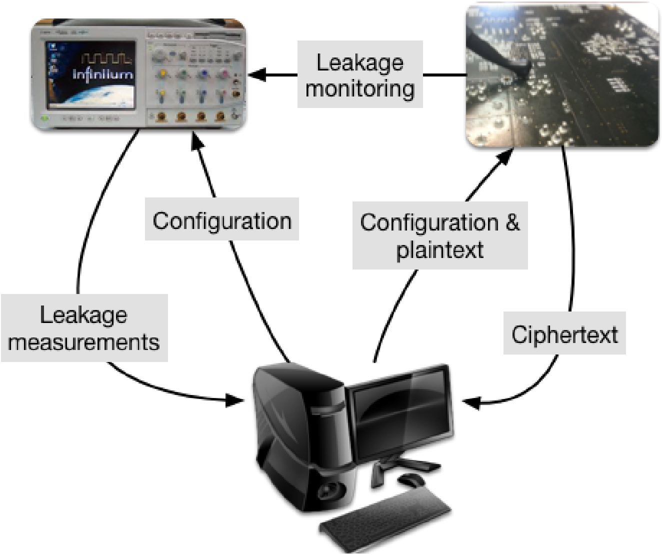
- embedded symmetric crypto on secure chips (e.g., AES-256)



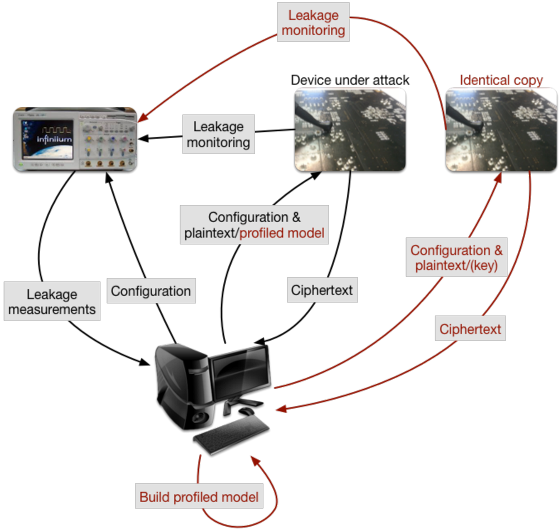
- the device *leaks* through a side channel



Acquisition Platform



Acquisition Platform (Profiled Scenario)



Mathematical Problem

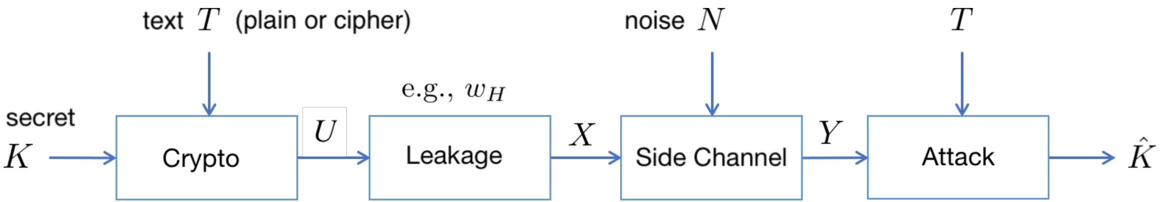
Can we derive

- the maximum attack success rate for a given number q of queries (traces)
- minimum # $q_{\min}$ of queries (traces) for a given level of attack success (e.g., 95%)

for

- any type of attack (DoM, DPA, CPA, LRA, MIA, KSA, other as-yet-unknown. . .)
- an omniscient/almighty (best possible) attacker
  - knows how the device leaks, everything except the secret
  - worst case for the defender

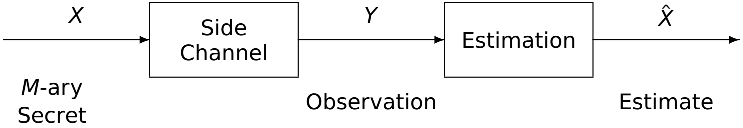
Formalization



Framework of **[Cherisey-Guilley-Rioul-Piantanida'19]**:

- AES-256 implementation with many (q) measurement traces
- Hamming weight leakage model $Y_i = w_H(S(T_i \oplus K)) + N_i \qquad (i = 1, 2, \ldots, q)$
- with or without countermeasures (shuffling, noising, masking)
- optimal attack

Bayesian Hypothesis Testing



Maximize success probability $\mathbb{P}_s = \mathbb{P}(\hat X = X)$ (minimize error probability $\mathbb{P}_e = 1 - \mathbb{P}_s$)

$$\begin{aligned}
\mathbb{P}(\hat X = X) &= \mathbb{E}\big(\mathbb{P}(\hat X = X|Y)\big) \\
&= \mathbb{E}\Big(\sum_x p(x|Y)\mathbb{P}(\hat X = x|Y)\Big) \qquad \text{since } X - Y - \hat X \text{ is Markov} \\
&\leq \mathbb{E}\big(\max_x p(x|Y)\big)
\end{aligned}$$

with equality if $\mathbb{P}(\hat X = x|Y) = 1$ for some x achieving $\max_x p(x|Y)$.

MAP (maximum a posteriori) rule

Maximum success $\boxed{\mathbb{P}_s(X|Y) = \mathbb{E}\big(\max_x p(x|Y)\big)}$ attained with $\hat X = \hat x(Y) = \arg\max_x p(x|Y)$.

MAP rule

- using disclosed measurements Y (output of a side channel):

$$\mathbb{P}_s(X|Y) = \mathbb{E}\Big(\max_x p(x|Y)\Big)$$

- prior belief ("blind", without access to the measurements):

$$\mathbb{P}_s(X) = \max_x p(x) = \max_x \mathbb{E}\big(p(x|Y)\big)$$

Theorem (Data Processing Inequality)

if $X - Y - Z$ is Markov,

$$\mathbb{P}_s(X|Y) \geq \mathbb{P}_s(X|Z)$$

In particular,

$$\boxed{\mathbb{P}_s(X|Y) \geq \mathbb{P}_s(X)}$$

Proof.

$\max \mathbb{E} \leq \mathbb{E} \max$. □

Guessing Entropy

In a game of "20 questions", what is the min average # questions before X is found?

- arbitrary questions: by dichotomy (probability 1/2-1/2): $H(X)$ (entropy)
- yes/no questions: by ranking (most probable first): $G(X)$ (guessing entropy)

- $\mathcal{G}(X)$ = number of successive guesses before secret X is found.
  Optimal strategy: $\mathcal{G}(X) = k$ guesses with probability $p_{(k)}$ (kth largest probability)
- guessing entropy $\boxed{G(X) = \min \mathbb{E}(\mathcal{G}(X)) = \sum_{k=1}^{M} k p_{(k)}}$ **[Massey'94]**
- with side information Y: $\boxed{G(X|Y) = \mathbb{E}_y G(X|Y=y)}$

Theorem (Data Processing Inequality)

if $X - Y - Z$ is Markov,

$$G(X|Y) \leq G(X|Z)$$

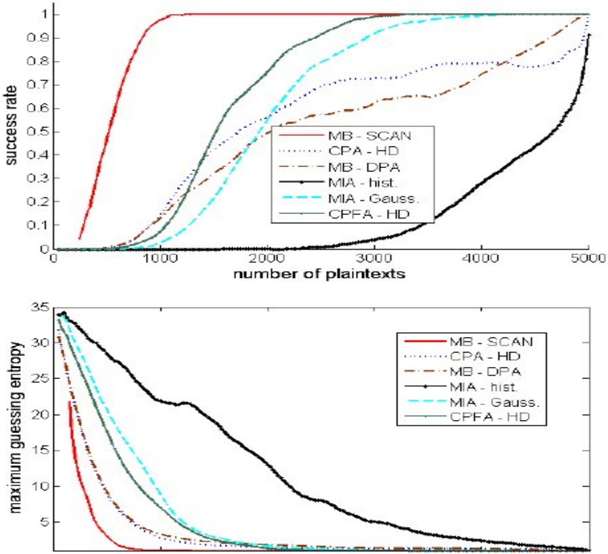
In particular,

$$\boxed{G(X|Y) \leq G(X)}$$

Proof.

W.l.o.g. $X = \mathcal{G}(X)$. Then $G(X|Y{=}y) \leq E(X|Y{=}y)$ so $G(X|Y) \leq \mathbb{E}(X) = G(X)$. □

Success Rate vs. Guessing Entropy



α-Entropy

- α-entropy: $\boxed{H_\alpha(X) = \frac{\alpha}{1-\alpha} \log \|p_X\|_\alpha}$ **[Rényi'61]**

  where "norm" $\|p\|_\alpha = \left(\int p^\alpha \,\mathrm{d}\mu\right)^{1/\alpha}$ is $\begin{cases} \text{convex (Minkowski) } \alpha > 1 \\ \text{concave (reverse Minkowski) } \alpha < 1 \end{cases}$

- conditional α-entropy: $\boxed{H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}_Y \|p_{X|Y}\|_\alpha}$ **[Arimoto'75]**

  (expectation *inside* the log)

Theorem (Data Processing Inequality)

if $X - Y - Z$ is Markov, $H_\alpha(X|Y) \leq H_\alpha(X|Z)$

In particular, $\boxed{H_\alpha(X|Y) \leq H_\alpha(X)}$

Proof.

$\mathbb{E}\|\cdot\| \underset{\alpha>1}{\overset{\alpha<1}{\lessgtr}} \|\mathbb{E}\cdot\|$. One recovers the usual DPI for the limit case $\alpha = 1$. □

MAP and min-Entropy, Guessing and $1/2$-entropy

$$
\begin{array}{ccccc}
H_0(X) = \log M & \geqslant \cdots \geqslant & H(X) = H_1(X) & \geqslant \cdots \geqslant & H_\infty(X) = \log \dfrac{1}{\mathbb{P}_s(X)} \\
\veebar & & \veebar & & \veebar \\
H_0(X|Y) = \log M & \geqslant \cdots \geq & H(X|Y) = H_1(X|Y) & \geqslant \cdots \geqslant & H_\infty(X|Y) = \log \dfrac{1}{\mathbb{P}_s(X|Y)}
\end{array}
$$

$\alpha \to 0$: Hartley's information theory

$\alpha \to 1$: Shannon's information theory

$\alpha \to +\infty$: estimation theory

Arikan's inequalities: **[Arikan'96]** useful for scalability **[Choudary17]**

$$
\begin{array}{ccccc}
H_{1/2}(X) - \log(1 + \ln M) & \leq & \log G(X) & \leq & H_{1/2}(X) \\
\veebar & & \veebar & & \veebar \\
H_{1/2}(X|Y) - \log(1 + \ln M) & \leq & \log G(X|Y) & \leq & H_{1/2}(X|Y)
\end{array}
$$

What is α-Information Theory?

- α-entropy $H_\alpha(X)$ or $H_\alpha(p)$
- α-conditional entropy $H_\alpha(X|Y)$
- α-divergence (relative entropy) $D_\alpha(p\|q)$
- α-information $I_\alpha(X;Y)$
- α-conditional information $I_\alpha(X;Y|Z)$

with interesting properties:

- consistency, nonnegativity, uniform expansion, relationships, ...
- conditioning reduces entropy
- data processing decreases information
- Fano's inequality

where $\alpha \in (0,1) \cup (1,+\infty)$ with limiting cases:

$$\begin{aligned} \alpha \to 0 &: \text{Hartley's information theory} \\ \alpha \to 1 &: \text{Shannon's information theory} \\ \alpha \to +\infty &: \text{estimation theory (MAP)} \end{aligned}$$

$D_\alpha(p\|q)$: α-Divergence

$$\boxed{D_\alpha(p\|q) = \frac{\alpha}{\alpha-1}\log(p\|q)_\alpha} \qquad \textbf{[Rényi'61]}.$$

where α-"product" $(p\|q)_\alpha = \left(\int p^\alpha q^{1-\alpha}\,\mathrm{d}\mu\right)^{1/\alpha}$

Properties:

- $D_\alpha(p\|q) \underset{\alpha\to 1}{\longrightarrow} D(p\|q)$ (Kullback-Leibler)
- binary expression $d_\alpha(p\|q) = \frac{1}{\alpha-1}\log\bigl(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}\bigr)$
- μ-independent! ☺ $p\,\mathrm{d}\mu = p'\,\mathrm{d}\mu$ & $q\,\mathrm{d}\mu = q'\,\mathrm{d}\mu \implies (p/q)^\alpha q\,\mathrm{d}\mu = (p'/q')^\alpha q'\,\mathrm{d}\mu'$
- nonnegative: $\boxed{D_\alpha(p\|q) \geq 0}$ since $(p\|q)^\alpha_\alpha \underset{\alpha>1}{\overset{\alpha<1}{\lesseqgtr}} \underbrace{(\int p)^\alpha (\int q)^{1-\alpha}}_{=1}$ $\begin{cases} \text{(Hölder) } \alpha < 1 \\ \text{(reverse Hölder) } \alpha > 1 \end{cases}$
- uniform expansion: if $u \equiv \frac{1}{M}$ then $(p\|u)_\alpha = M^{\frac{\alpha-1}{\alpha}}\|p\|_\alpha$ and $\boxed{D_\alpha(p\|u) = \log M - H_\alpha(p)}$
  hence $H_\alpha(X) \leq \log M$ = max α-entropy for uniform

$D_\alpha(p\|q)$: α-Divergence

- data processing inequality:

by the "golden formula" $\boxed{(p_{XY}\|q_{XY})_\alpha = \big(p_X\,(p_{Y|X}\|q_{Y|X})_\alpha \,\big\|\, q_X\big)_\alpha} \overset{\alpha<1}{\underset{\alpha>1}{\lesseqgtr}} (p_X\|q_X)_\alpha$

one has $D_\alpha(p_{XY}\|q_{XY}) \geq D_\alpha(p_X\|q_X)$ with equality if $p_{Y|X} = q_{Y|X}$. Therefore,

$$\text{if} \quad \begin{cases} p_X \to \boxed{p_{Y|X}} \to p_Y \\ q_X \to \boxed{p_{Y|X}} \to q_Y \end{cases} \quad \text{then} \quad \boxed{D_\alpha(p_Y\|q_Y) \leq D_\alpha(p_X\|q_X)}$$

Example: $X \to \boxed{1_A} \to Y$

$$D_\alpha(p\|q) \geq d_\alpha(p_A\|q_A) \text{ where } p_A = \mathbb{P}(X\in A),\ q_A = \mathbb{Q}(X \in A).$$

Example: binary channel $X \to \boxed{p_{Y|X}} \to Y$

$$d_\alpha(p\|r) \geq d_\alpha(p\|q) \quad \text{and} \quad d_\alpha(p\|r) \geq d_\alpha(q\|r)$$

for any p, q, r in that order ($p \leq q \leq r$ or $p \geq q \geq r$).

$I_\alpha(X;Y)$: α-Information

Consider $D_\alpha(p_{X|Y=y}\|p_X) = \frac{\alpha}{\alpha-1}\log(p_{X|Y=y}\|p_X)_\alpha$ and take the expectation over Y inside the logarithm:

$$\boxed{I_\alpha(X;Y) = \frac{\alpha}{\alpha-1}\log \mathbb{E}_Y(p_{X|Y}\|p_X)_\alpha}$$

that is, $I_\alpha(X;Y) = \frac{\alpha}{\alpha-1}\log\int_{\mathcal{Y}}\Big(\int_{\mathcal{X}} p(x)p^\alpha(y|x)\,\mathrm{d}\mu(x)\Big)^{1/\alpha}\mathrm{d}\mu(y) =$
$\frac{\alpha}{\alpha-1}\log\int_{\mathcal{Y}} p(y)\Big(\int_{\mathcal{X}} p^\alpha(x|y)p^{1-\alpha}(x)\,\mathrm{d}\mu(x)\Big)^{1/\alpha}\mathrm{d}\mu(y)$ **[Sibson'69]**

Properties:

- μ-independent! ☺
- uniform expansion: if $U\sim\mathcal{U}(M)$ then $I_\alpha(U;Y) = \log M - H_\alpha(U|Y)$
  but $I_\alpha(X;Y) \neq H_\alpha(X) - H_\alpha(X|Y)$ in general
- not mutual: $I_\alpha(X;Y) \neq I_\alpha(Y;X)$ in general

$I_\alpha(X;Y)$: α-Information

Properties (cont'd)

- Sibson's identity (golden formula): by the golden formula $(p_{XY}\|p_Xq_Y)_\alpha = \big(\underbrace{p_Y\,(p_{X|Y}\|p_X)_\alpha}_{\propto\, q_Y^*} \,\big\|\, q_Y\big)_\alpha$, we have

$$\boxed{D_\alpha(p_{X,Y}\|p_Xq_Y) = D_\alpha(q_Y^*\|q_Y) + I_\alpha(X;Y)}$$

- in particular $I_\alpha(X;Y) = \min_{q_Y} D_\alpha(p_{X,Y}\|p_Xq_Y) \geq 0$ (nonnegative) and $= 0$ iff $X \perp\!\!\!\perp Y$
- data processing inequality: if $W - X - Y - Z$ is Markov, by the data processing inequality for α-divergence: $D_\alpha(p_{XY}\|p_Xq_Y) \geq D_\alpha(p_{WZ}\|p_Wq_Z)$, hence

$$\boxed{I_\alpha(X;Y) \geq I_\alpha(W;Z)}$$

$I_\alpha(X;Y|Z)$: Conditional α-Information

Consider $I_\alpha(X;Y|Z=z) = \frac{\alpha}{\alpha-1} \log \mathbb{E}_Y (p_{X|Y,Z=z} \| p_{X|Z=z} \rangle_\alpha$ and take the expectation over Z inside the logarithm:

$$I_\alpha(X;Y|Z) = \frac{\alpha}{\alpha-1} \log \mathbb{E}_{YZ} (p_{X|Y,Z} \| p_{X|Z} \rangle_\alpha$$

[Liu,Cheng,Guilley,Rioul'21]

- consistent: $I_\alpha(X;Y|0) = I_\alpha(X;Y)$
- uniform expansion: $I_\alpha(U;Y|Z) = \log M - H_\alpha(U|YZ)$
- golden formula: $D_\alpha(p_{XYZ} \| p_{X|Z} q_{YZ}) = D_\alpha(q^*_{YZ} \| q_{YZ}) + I_\alpha(X;Y|Z)$ hence $I_\alpha(X;Y|Z) \geq 0$ (nonnegative) and $= 0$ iff $X - Z - Y$

Other definitions

- α-entropy:
  - Tsallis **[Havrda-Charvát'67]** $\dfrac{1 - e^{(1-\alpha)H_\alpha(X)}}{\alpha - 1}$
    not even constant in α for uniform X ☹
- conditional α-entropy:
  - $\mathbb{E}_Y H_\alpha(X|Y=y)$ **[Cachin'97]**; not monotonic ☹
  - $H_\alpha(X,Y) - H_\alpha(Y)$ **[Golshani+al'09]**; not monotonic ☹
  - $\frac{1}{1-\alpha} \log \mathbb{E}_Y \|p_{X|Y=y}\|_\alpha^\alpha$ **[Hayashi'11]**; no chain rule ☹
  - $-\log \mathbb{E}_y \|p_{X|Y=y}\|_\alpha^{\frac{\alpha}{\alpha-1}}$ **[Fehr-Berens'14]**; no chain rule ☹

Other definitions

- α-information:
  - $H_\alpha(X) - H_\alpha(X|Y)$ **[Arimoto'75]** no data processing inequality ☹
  - $D_\alpha(p_{XY} \| p_X p_Y)$; no uniform expansion ☹
  - $\min_{q_Y} \mathbb{E}\, D_\alpha(p_{Y|X} \| q_Y)$ **[Augustin'78][Csiszár'95]** no uniform expansion, no data processing inequality ☹
  - $\min_{q_X, q_Y} D_\alpha(p_{XY} \| q_X q_Y)$ **[Lapidoth-Pfister'16]** (symmetric) not even closed-form ☹
  - etc.
- conditional α-information:
  - $D_\alpha(p_{XYZ} \| p_{X|Z} p_{Y|Z} P_Z)$ not consistent ☹
  - $\min\limits_{q_{Y|Z}} D_\alpha(p_{XYZ} \| p_{X|Z} q_{Y|Z} p_Z)$ **[Tomamichel-Hayashi'18]** no unif. expansion ☹
  - $\min\limits_{q_Z} D_\alpha(p_{XYZ} \| p_{X|Z} p_{Y|Z} q_Z)$ **[Esposito+al'21]** not consistent ☹
  - etc.

α-Fano Inequality for α-Information

$X - Y - \hat{X}$ with M-ary X, probability of success $\mathbb{P}_s = \mathbb{P}(\hat{X} = X)$

- X is a sensitive data (depending on a secret);
- $P_{Y|X}$ is a "side-channel" through which information leaks
- Y is disclosed to the attacker (measurements by probes/sniffers...)
- $P_{\hat{X}|Y}$ is the attack (MAP rule maximizes probability of success)

$$I_\alpha(X;Y) \underset{DPI}{\geq} I_\alpha(X,\hat{X}) = D_\alpha(p_{X,\hat{X}} \| p_X q^*_{\hat{X}}) \underset{DPI}{\geq} d_\alpha(\mathbb{P}_s(X|Y) \| \mathbb{P}'_s) \underset{dpi}{\geq} d_\alpha(\mathbb{P}_s(X|Y) \| \mathbb{P}_s(X))$$

where $\mathbb{P}'_s = \sum_x p_X(x) q^*_{\hat{X}}(x) \leq \max_x p_X(x) = \mathbb{P}_s(X)$.

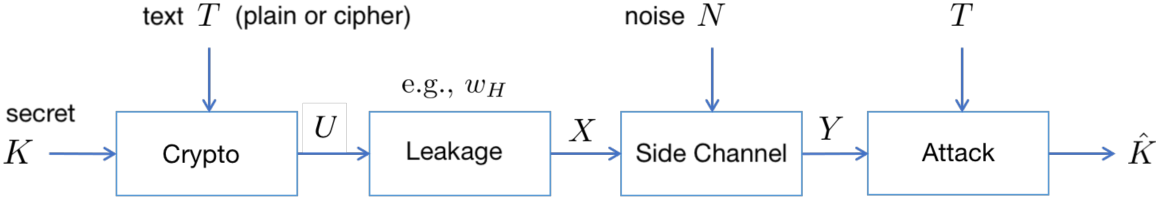
α-Fano's Inequality [Rioul'21]

$$I_\alpha(X;Y) \geq d_\alpha\big(\mathbb{P}_s(X|Y) \,\|\, \mathbb{P}_s(X)\big)$$

generalizes **[HanVerdú'94]** ($\alpha = 1$)

$\Longrightarrow$ implicit upper bound on $\mathbb{P}_s(X|Y)$ as a function of α-information.

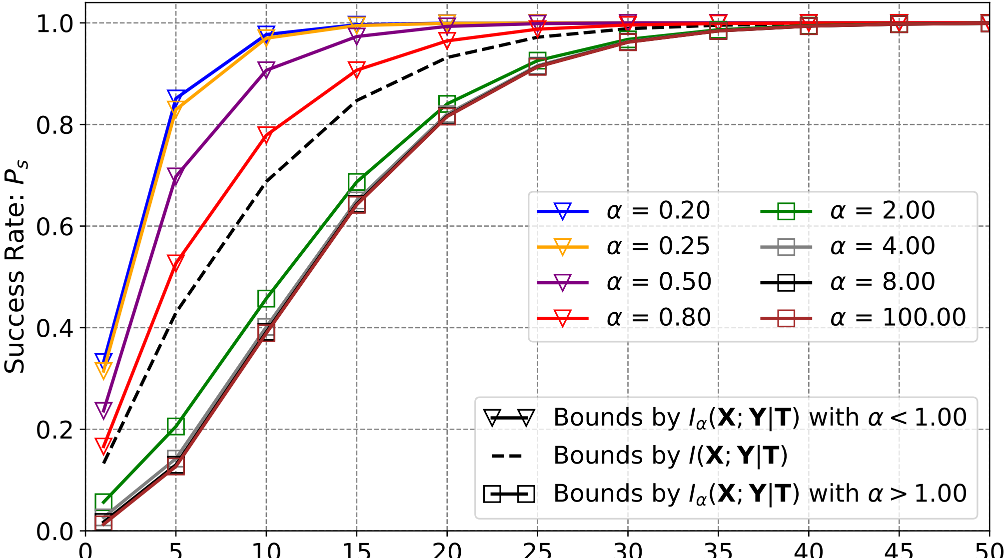
Application to Side-Channel Analysis



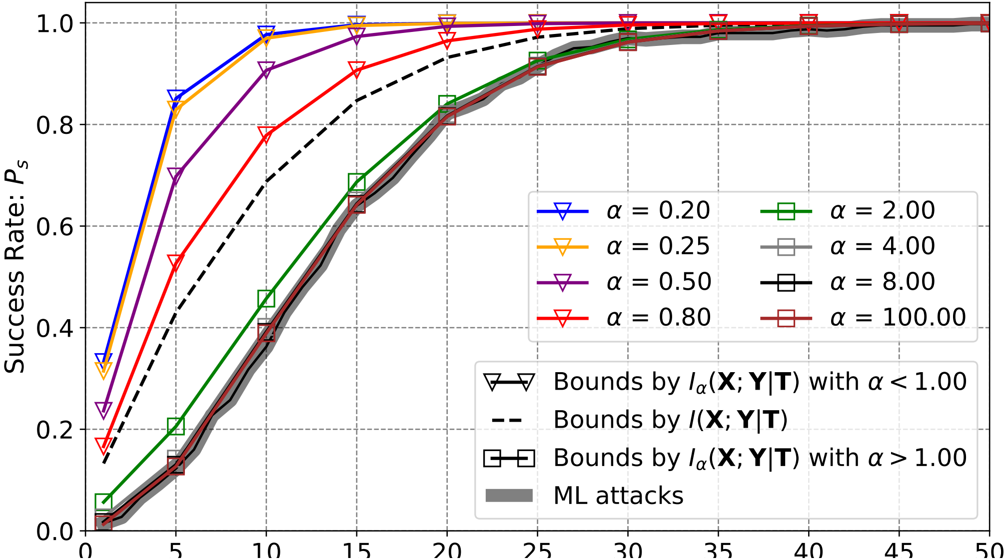
Framework of **[Cherisey-Guilley-Rioul-Piantanida’19]**:

- AES-256 implementation with many (q) measurement traces
- Hamming weight leakage model $Y_i = w_H(S(T_i \oplus K)) + N_i \qquad (i = 1, 2, \ldots, q)$
- $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) \geq d_\alpha\left(\mathbb{P}_s \,\|\, \frac{1}{M}\right)$ by the main theorem applied to $K - \mathbf{X} - \mathbf{Y}$
- Monte-Carlo simulation to compute $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$
- upper bound success rate $\mathbb{P}_s$ as a function of q
- lower bound # traces $q_{\min}$ needed to achieve a given success $\mathbb{P}_s$
- compare to optimal (maximum likelihood) attacks giving $\mathbb{P}_s(K|Y)$

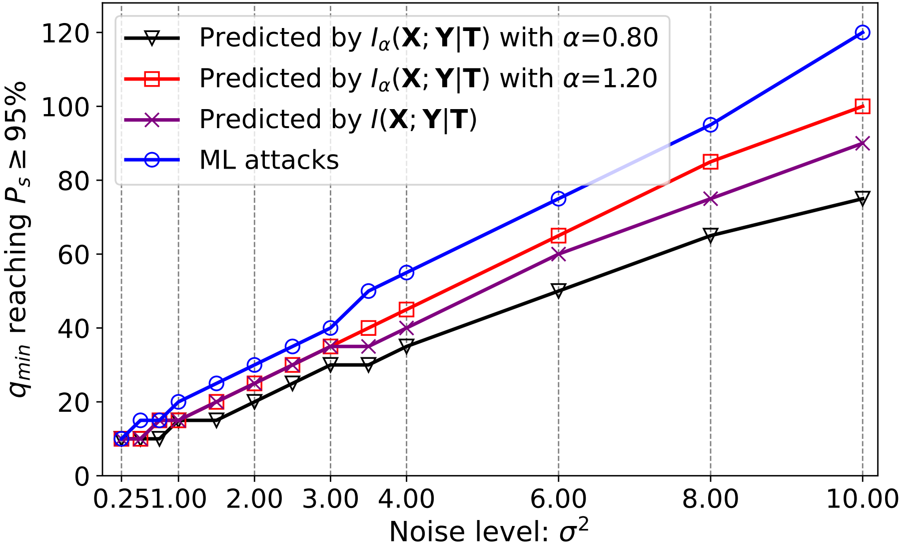
Upper Bounds on Success Rate $\mathbb{P}_s$



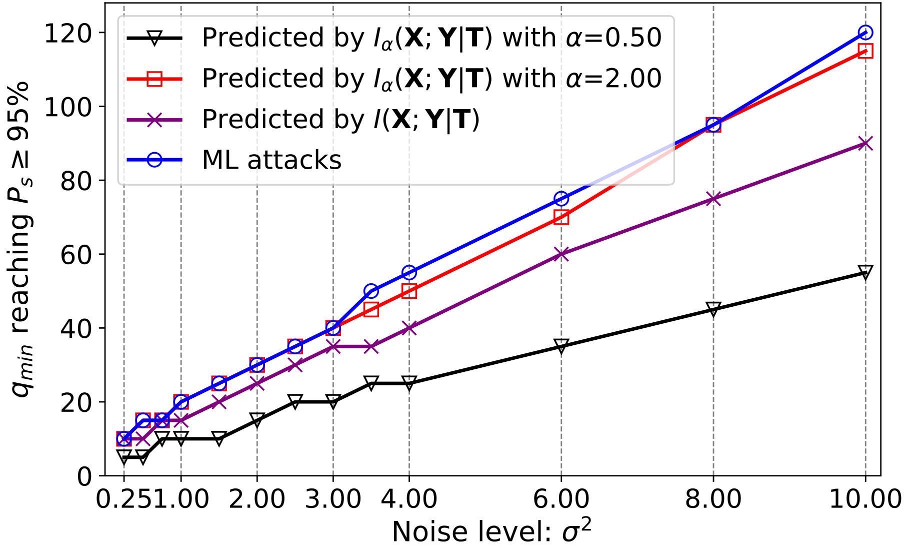
Upper Bounds on Success Rate $\mathbb{P}_s$



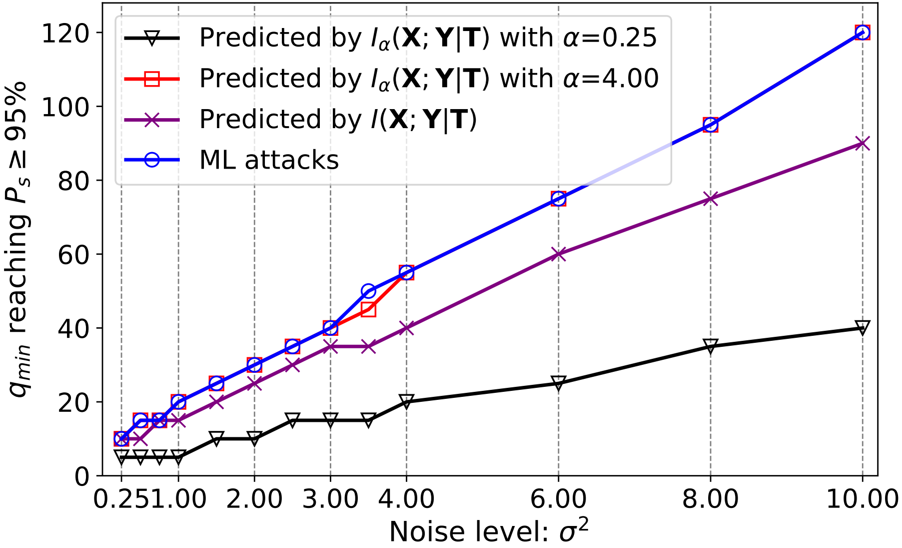
Lower Bounds on # of Traces to Achieve 95% Success



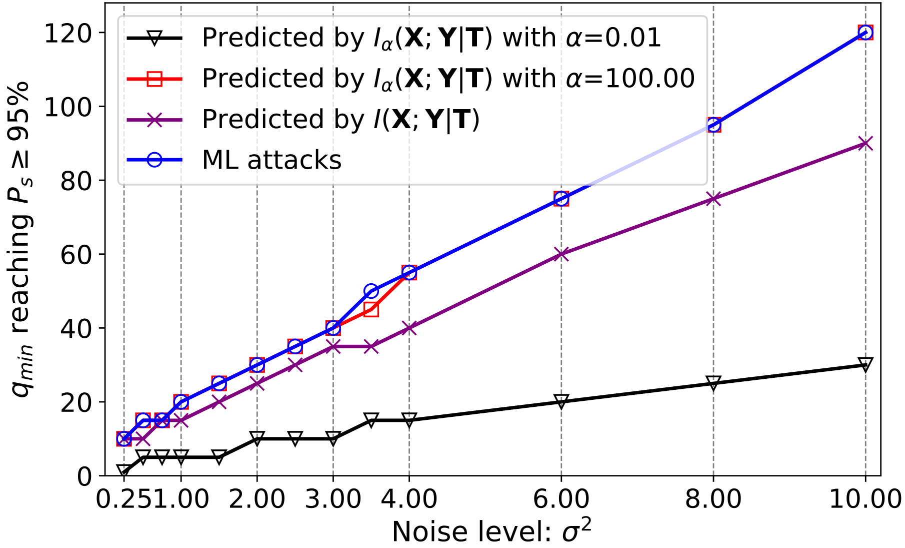
Lower Bounds on # of Traces to Achieve 95% Success



Lower Bounds on # of Traces to Achieve 95% Success



Lower Bounds on # of Traces to Achieve 95% Success



Information Leakage and Side-Channel Attacks

Fuites d’information et attaques par canaux cachés

Merci !

Olivier Rioul

Télécom Paris, Institut Polytechnique de Paris, France

<olivier.rioul@telecom-paris.fr>